Mixed results for gendered patterns in confidence of team success and collective efficacy

Abstract

Gendered differences in academic confidence and self-efficacy between men and women are well-documented. In STEM fields and specifically in engineering, such differences have important consequences in that students low on these constructs are often more prone to leave their degree programs. While this evidence base is fairly established, less is known about the extent to which men and women show differences in confidence of *team* success, or collective efficacy, which may also be consequential in decisions to join and persist in design team experiences, or even to stay in or leave an engineering major, especially for first-year students. In this analysis, we quantitatively investigated gendered differences in confidence of team success and collective efficacy among first-year engineering students working on semester-long design projects in stable teams. Using a software tool built to support equitable teamwork, survey data on team confidence and collective efficacy was collected for these engineering students as well as for students in other courses for the sake of comparison. Three hierarchical linear models were fit to the data from 1,806 students across 31 unique course/term combinations. The results were mixed. In two of these analyses, we identified significant interactions between gender and team confidence. Specifically, men generally reported higher team confidence scores than women throughout the term with women eventually catching up, and team confidence ratings increased for men but not women following a lesson on imposter syndrome. No gendered differences were observed with respect to a collective efficacy scale administered near the middle and end of the term, however. In all cases, the results were consistent across course type (engineering, business, and others).

Keywords: collective efficacy, engineering, first-year students, gender, team confidence, team success

Introduction

Gendered differences in academic confidence and self-efficacy

The confidence with which students go about their academic endeavors is clearly related to important educational outcomes like learning and persistence. As a general concept, academic confidence is based on feelings, emotions, and past experiences as well as on the judgements of oneself and others about one's abilities to execute behaviors towards different goal outcomes [1], [2]. Related to academic confidence, self-efficacy is the context-specific belief in one's capabilities to complete a given task [3]. As a foundational idea in educational psychology, self-efficacy has been widely studied and shown to positively impact academic performance [4], [5]. The positive impact of higher self-efficacy has been shown to remain after controlling for variation in several covariates, such as socioeconomic status and measures of prior performance [6]. Unlike personality traits, which tend to be more stable, self-efficacy in a particular domain is a unique contributor to student success that can be affected with targeted interventions. Further, self-efficacy has a reciprocal relationship with performance, meaning positive academic performance can subsequently and uniquely improve self-efficacy [7]. As a result, there is an opportunity to address attrition in a discipline by intentionally working to improve students' self-efficacy in relevant domains.

Academic confidence is usually considered to be more domain general while self-efficacy is considered more task-specific [4]. Indeed, the two constructs have been shown to load as separate dimensions [8], though in this case, confidence in the correctness of assessment responses (that is, a specific scenario) was being considered. In any case, in this study, we are actually not so concerned with the precise demarcations between confidence, self-efficacy, and related ideas like self-concept. Rather, we sought to carry out a basic exploration of these ideas at the *team*, rather than individual, level. In particular, we intended to explore the extent to which confidence in team performance and success does, or does not, exhibit gendered patterns for undergraduate students working on semester-long course project teams.

Gendered differences in confidence and self-efficacy constructs have been observed for many years across domains and disciplines [9]. For example, D'Lima and colleagues [10] showed that across several large introductory courses at their university, first-year male undergraduate students exhibited greater self-efficacy with respect to their academic abilities than female students both at the beginning and end of their first term of college, though self-efficacy increased significantly for both groups over the term. Huang's [11] meta-analysis of gendered differences across nearly 200 studies showed female students having greater self-efficacy in language arts and male students having greater self-efficacy in social sciences, computer science, and mathematics. At the extreme, of course, overconfidence is a liability: Johnson [12], for example, showed that for male students, confidence in academic ability was actually negatively correlated with course grades whereas female students did not show the same pattern. Several studies have also reported a lack of gendered patterns in self-efficacy, perhaps in part due to "the microanalytic nature of efficacy constructs" [13]. In other words, self-efficacy is related to gender, but because self-efficacy can be defined at the task level and different domains have different biases, we do not see universally true patterns.

Gendered differences in academic confidence and self-efficacy specifically in STEM domains

Our work is particularly concerned with building foundational knowledge and instructional tools that support institutionalizing transformed teaching and learning practices across undergraduate science, technology, engineering, and mathematics (STEM) courses. In general, STEM disciplines have maintained exclusionary behaviors (like the common practice of providing low average grades in foundational courses [14]) that have contributed to and exacerbated gendered differences in undergraduate student persistence and degree attainment for decades [15]. Each discipline certainly has its own specific trajectory and stories [16], but supporting adequate gender representation and student thriving across STEM disciplines remains the subject of much research and many funding programs. Recent data shows that the lack of female student representation remains most stark in computer science, engineering, and physics (~20%) while greater numerical parity is observed in biology, chemistry, and mathematics [17]. In our own College of Engineering, a key context for this study, 28% of degrees are currently awarded to women [18] compared to the national average of 22% [19]. While the proportion of women enrolled has slowly increased in recent years, challenges remain with respect to enrollment and climate [20].

Across STEM fields, the general finding across a significant number of studies has been that men have higher academic confidence and self-efficacy than women [21]–[24]. Specific studies with undergraduate mathematics [4], physics [25], and engineering [2] students have borne out this result, showing relationships between these differences and subsequent academic success. Studies in engineering on the interplay between gender and the ever-important social constructs of race and ethnicity have further shown that white men in particular tend to demonstrate the highest levels of confidence and self-efficacy [26], [27]. Of course, it is important to recognize that within the large body of STEM-specific research on these constructs, some variation naturally exists. Ross and colleagues [28], for example, described conflicting bodies of research about whether the gendered self-efficacy gap in mathematics has been maintained. Some studies in engineering have similarly reported a lack of difference in self-efficacy by gender [29], [30] though Mamaril and colleagues [30] reported the exception that men reported marginally higher self-efficacy than women for "tinkering" or the manual manipulation of tools. In any case, differences in confidence and self-efficacy have been linked to the serious consequences of leaving particular threads of coursework and degree programs altogether [15], [31], [32] and, as such, are worthy of continued investigation.

Gendered differences in team confidence and collective efficacy

Now, we turn to the heart of this study—extending these questions about gendered differences in confidence and efficacy to the setting of project teams in undergraduate courses.

Bandura's classic social cognitive theory (SCT; [1]) partitions the determinants of human behavior into three separate yet interdependent areas: behavioral factors like the ability for students to self-regulate their actions towards positive outcomes, environmental or situational factors like the the classroom context, and personal factors (sometimes alternatively referred to as cognitive factors) like attitudes and values. SCT describes learning as occurring through observation of others in a social context and, importantly, an "agentic" perspective that relies heavily on self-efficacy, one of the key theoretical constructs supporting SCT [33]. The importance of self-efficacy in motivating an individual's behaviors and actions towards their goals can hardly be overstated, and yet the broader notions of team confidence and collective efficacy are very useful for understanding how an interdependent group of people behaves in reaching (or not reaching) their collective goals [4], [34], [35].

Collective efficacy is a form of agency that emerges for an interactive system, like a team [36], and it has been shown to better predict group performance as compared with individual team members' self-efficacy [37]. As a framework for understanding if and how a team can achieve their goal, collective efficacy begins to touch the realms of culture, organizational leadership and power, social systems, and systems thinking. In short, do the team members trust the whole to operate effectively, with purpose, towards a useful and desired end? The idea of collective efficacy has been employed effectively across a wide range of disciplines from psychology and education to business, management, criminology, urban sociology, and sports, providing a useful lens, for example, to help explain differences between highly interdependent teams (as in basketball and soccer) versus more individual sports teams (as with gymnastics and swimming) [38]–[41]. Notably, because collective efficacy is not simply the sum of each individual's appraisal of their ability to do their own job, measuring this construct can be somewhat difficult [33].

Prior work investigating gendered patterns in team confidence and collective efficacy shows conflicting results. Some studies show linkages between gender, collective efficacy, and team performance. For example, Niler and colleagues [42] recently showed that for undergraduate women working on a semester-long course project, having a larger percentage of women on their team was associated with both increased collective efficacy and identification with the team. Further, both of these constructs—collective efficacy and team identification—acted as mediators for the effect of percent of women on the team and team performance. None of these results were observed for men in their sample. In contrast, students' beliefs about their collective efficacy as a class (e.g., for learning and maintaining a supportive classroom environment) have been shown to be significantly related to individual students' grades for Portuguese boys but not girls at the secondary level [13].

Still other work, notably conducted in settings outside of the United States, has found gender to be altogether unrelated to collective efficacy. Forslund Frykedal and colleagues [43], for example, showed that gender was a non-significant factor in predicting collective efficacy among Swedish students working on teams in school years 5 and 8; here, self-efficacy and interdependence accounted for most of the variance in collective efficacy. Within a non-educational setting of sales representative teams at a Taiwanese financial institution, researchers also found no effect of gender diversity on group cohesion or collective efficacy [44]. Overall, the limited number of studies, and particularly the limited number of studies specifically within the undergraduate STEM education context, provides justification for our interest in gendered notions of team confidence and collective efficacy.

Research questions

Thus, our goal here was to understand if gendered patterns are evident in undergraduate student teams working on course projects with respect to confidence of team success or collective efficacy. Towards this end, we pursued the notion of gendered patterns with respect to the following research questions: (RQ1) How do team confidence ratings change over the course of the term? (RQ2) How do collective efficacy ratings change over the course of the term? And (RQ3) To what extent does exposure to a lesson on imposter syndrome impact students' assessment of confidence in their team?

Context

Data was collected at our large, public, primarily residential four-year university using an in-house developed software tool for teaming and teamwork called Tandem. Tandem serves as a mechanism for surveying students with respect to teamwork throughout the term and acts as a platform for lesson and reflection content about teamwork topics (these and other functions of Tandem, like providing feedback to instructors, are described in [45] in greater depth).

Importantly, Tandem is built on a foundational engine that supports tailoring content to students based on their responses to surveys, academic history, and in-course engagement and outcomes. For example, a student who notes on the beginning-of-term survey that they are usually reluctant to share their ideas and preferences in groups would see content in Tandem messages and lessons throughout the term that broadly encourages them to speak up. Further, the Tandem content is generally written in a supportive, motivational interviewing style [46]. That is, instead of directly telling a reluctant student that they should speak up more, Tandem might ask the student what they'd say to a friend who is reluctant to speak up.

Tandem was first launched in the academic year 2018-2019 in a 100-level engineering design course. In every term since, and notably throughout the COVID-19 pandemic, new faculty partners in additional courses have begun to use Tandem. In the current academic year (2021-2022), Tandem has been used in 11 unique courses across engineering, business, history, information, kinesiology, public health and other disciplines largely at the undergraduate level, covering approximately 3,500 student enrollments. Tandem is also used to support students in a few small-scale co-curricular learning experiences, such as an interdisciplinary master's level fellowship program for students interested in sustainability.

Methods

Course groups

In the current study, which is considered exempt from review by our Institutional Review Board, we considered all the courses (and none of the co-curricular learning experiences) in which Tandem was used during the six terms between Winter 2019 and Fall 2021 (excluding summer terms). Each course/term combination was considered separately and we formed three groups for the sake of comparison.

The first group (engineering) denotes all instances of the cornerstone first-year engineering design course at our university. This group was of primary interest for this study because gendered differences in confidence are particularly stark in STEM courses and because the first two years of STEM courses have been found to be the most impactful on student retention [15], [47].

The second group (business) denotes all instances of a lower-division business course on leadership and society targeted toward second-year students. This course is designed to expose tensions and opportunities between business and broader societal issues; the use of Tandem in this course is described in greater depth elsewhere [48]. We selected this course as the key comparator because, after engineering, it is the largest-enrollment lower-division course that makes use of Tandem. In contrast with the smaller, upper-division courses making use of Tandem, students in these engineering and business courses are most likely to be using Tandem for the first time.

The third group (other) denotes all instances of all other courses that used Tandem between Winter 2019 and Fall 2021. We note that Tandem is used in a few courses that run over two semesters, such as a senior-level capstone course in information. In these cases, data from both semesters were included.

Data collection

Demographic data were self-reported by the students on the Tandem beginning-of-term survey. Students were asked to select the gender with which they identify and had the option to select female (43.3%), male (55.4%), non-binary (0.7%), prefer not to disclose (0.4%), or prefer to self-describe with a write-in field (0.1%); only the female and male categories had sufficient numbers of student to be retained in the statistical analyses. Importantly, while we recognize that the ability for students to self-report these data is valuable, as most data recorded at the institutional level is limited in this regard, unfortunately the stem and response options here errantly conflate gender and sex [49]. This issue represents an important methodological limitation that is now being addressed for future Tandem implementations. With full recognition of this limitation, for the sake of consistency, herein we use language based on the construct of gender rather than assigned sex.

For each course, three sets of data were collected from the Tandem database:

Team confidence data. First, team checks are short, frequent surveys that students respond to in Tandem as a reflection about how their team is functioning; rating the team as a whole, rather than rating each individual separately, is a key emphasis of the survey. The team check has five closed-ended questions, each recorded on a Likert scale. We collected the data for the particular team check question about team confidence, for which the end anchors are "I worry we won't do well on this project" and "We're definitely going to do well on this project".

Importantly, the number of points on the Likert scale used for team checks has varied. In the vast majority of Tandem courses, a 9-point scale has been used. In Fall 2021, however, about half of Tandem courses used a 5-point scale. Historically, students have been reluctant to use the lower end of the team check scale, thus this change was temporarily implemented to test if it would encourage students to use responses at and below the neutral value. This variation between the 5-point and 9-point scales is accounted for in the analyses.

Further, team checks are usually administered on a biweekly basis throughout the term after project teams have been established, typically yielding around five time points of data in each semester. Because the exact team check schedule varies for each course (and the exact date on which the student completes the team check is uncontrolled), the timing of the team checks was unbalanced. This imbalance in measurement schedule is also controlled via random effects of time for students and courses.

Collective efficacy data. Second, we collected student responses to a collective efficacy scale [50] that students responded to twice during each term (Table 1). This scale was administered as part of a midterm feedback survey, usually occurring at the middle of the term, and an end-of-term survey, both of which function as opportunities for peer- and self-evaluation. Because these items were added to Tandem effective Winter 2021, only two semesters worth of data, and thus relatively few observations (figures are provided in the Data analysis subsection), were available for analysis.

Table 1. The collective efficacy (CE) items recorded on a 9-point scale with end anchors “not at all confident” (1) and “completely confident” (9). The stem on the midterm survey was “How confident are you that your team could…” and the stem on the end-of-term survey was “Please rate your agreement with the following statements. Our team was able to...”.
CE1: Reach agreement about what needs to get done at each meeting CE2: Find ways to bridge individual differences (e.g., in age, major, or personality) between team members CE3: Assist members who are having difficulty with certain tasks CE4: Develop a workable project in a reasonable amount of time CE5: Communicate well with one another despite differences in cultural background CE6: Adapt to changes in group tasks or goals CE7: Work well together even in challenging situations CE8: Deal with feedback or criticism from the course instructor CE9: Find ways to capitalize on the strengths of each member

Imposter syndrome lesson data. Finally, we collected student response data to the particular Tandem lesson about imposter syndrome [51]. The lesson, akin to a blog post or short news article, describes imposter syndrome, provides information about how common it is, defines five different types of imposter syndrome, and offers suggestions for how to deal with and ultimately overcome it. The lesson concludes with a short activity for students that asks them to identify the type of imposter syndrome most relevant to them (closed-ended question) as well as to answer a series of open-ended reflection questions. As with all the Tandem lessons, the timing of when the imposter syndrome lesson is presented to students varies based on instructor implementation of the tool, and the specific language in the lesson is tailored based on other data, in particular students’ responses to the Tandem beginning-of-term survey.

Data analysis

Three separate hierarchical linear models were estimated to predict student and course variation in two outcomes (team confidence and collective efficacy). Maximum likelihood estimation via SAS PROC MIXED with Satterthwaite denominator degrees of freedom was used to accommodate incomplete (e.g., missing) and unbalanced (e.g., individually-varying intervals between occasions of measurement) data and provide unbiased population estimates under the assumption that the data are missing at random; model syntax is available from the authors upon request.

The first analysis examined change in 13,356 team confidence ratings over time from 1,806 students nested within 31 courses (that is, course/term combinations), where time (i.e., the ordinal sequence of student team checks in a course; $M = 4.4$, $SD = 2.5$, $Max = 10$) was modeled as a quadratic function. In a quadratic model of change over time, the intercept is the predicted outcome whenever time = 0, the linear effect of time is the instantaneous linear rate of change in the outcome per unit time whenever time = 0 (i.e., the slope of the tangent line to the curve at that point in time), and twice the quadratic effect of time is how the linear effect of time changes per unit time (i.e., rate of acceleration or deceleration, which is not conditional on time = 0 when it is the highest-order polynomial term in the model). The second analysis examined change in 1,390 collective efficacy ratings from 711 students nested within 15 courses surveyed at about the midpoint and endpoint of a course (pretest/posttest design). The third analysis examined the effect of the imposter syndrome lesson on 3,018 team confidence ratings from the same 1,806 students within 31 courses as in the first analysis by comparing team confidence ratings in the team checks immediately preceding and following the lesson (pretest/posttest design). In each analysis, effects of gender (woman or man), course (engineering, business, or other), and their interaction were examined. The first and third analyses additionally considered the size of the team confidence Likert scale (5-point or 9-point).

For each analysis, model building began with an empty model and proceeded to add fixed and random effects starting with lower-level predictors (i.e., time-, student-, and then course-level predictors). Random effects were tested with *chi-square deviance tests* which compare the fit of models before and after adding random effects. Each model contained random intercepts and time slopes for students and courses (ps $< .001$) in order to control for variability between students and between courses in characteristics (intercepts) and change (time slopes). Fixed effects were tested with F-tests or *Wald t* tests where appropriate. Effect sizes were computed as *pseudo-*R^2, which describes the proportion of random effect variance reduced after adding predictors. Fixed effect patterns are described below.

Results

Analysis 1: Men report more team confidence than women early in the term, and women eventually catch up

To examine change in team confidence ratings over time (where time equals ordinal team check within a course; centered at the first team check), we began by inspecting the raw unstructured time trajectories (i.e., observed data with no random effects). As shown in the left panel of Figure 1, the overall trajectory appears to show quadratic change, which was supported by a significant quadratic effect of time ($b = -.0034$, $SE = .0001$, $p < .001$, *pseudo-*$R^2 = 2.35\%$). Accounting for 5-point and 9-point Likert items indicated a significant linear time x Likert item interaction, $F(1, 10960) = 7.14$, $p = .008$, *pseudo-*$R^2 = 4.39\%$, as well as a trending quadratic time x Likert item interaction, $F(1, 9923) = 2.56$, $p = .109$, *pseudo-*$R^2 = 0.64\%$. Therefore, to align the scales of the items, we normalized both by dividing every response by the maximum response of 9.

Importantly, there were significant linear (time = 0) and quadratic time interactions with gender (linear: $F(1, 9835) = 4.08$, $p = .044$, *pseudo-*$R^2 = 1.05\%$; quadratic: $F(1, 10820) = 5.02$, $p = .025$, *pseudo-*$R^2 = 4.95\%$). As shown in the right panel of Figure 1, the linear effect of time was significantly more positive for men ($b = .0128$, $SE = .0019$, $p < .001$) than women ($b = .0067$, $SE = .0023$). In other words, at the first team check, men's team confidence is predicted to increase more than women's. The quadratic effect of time, however, was significantly more negative for men ($b = -.0012$, $SE = .00029$, $p < .001$) than women ($b = -.0002$, $SE = .00032$, $p = .517$), such that the increase in confidence for men lessened over time more so than for women (in fact, the lessening of increase in confidence for women was not significant). The net result of these effects can be summarized as men reporting more team confidence than women early in the term, but this gradually changes to the point that women are predicted to eventually report more confidence than men.

The pattern of change over time for men and women did not vary significantly across the engineering, business, or other courses (linear time x gender x course: $F(2, 9945) < 1$, $p = .39$, $pseudo\text{-}R^2 = 0.97\%$; quadratic time x gender x course: $F(2, 10894) = 1.34$, $p = .26$, $pseudo\text{-}R^2 = 0.52\%$), or across the 5-point and 9-point Likert items (linear time x gender x Likert item: $F(1, 11001) < 1$, $p = .41$, $pseudo\text{-}R^2 = 0.51\%$; quadratic time x gender x Likert item: $F(1, 10098) < 1$, $p = .556$, $pseudo\text{-}R^2 = 0.01\%$). No other effects were significant either (*p*s > .33).

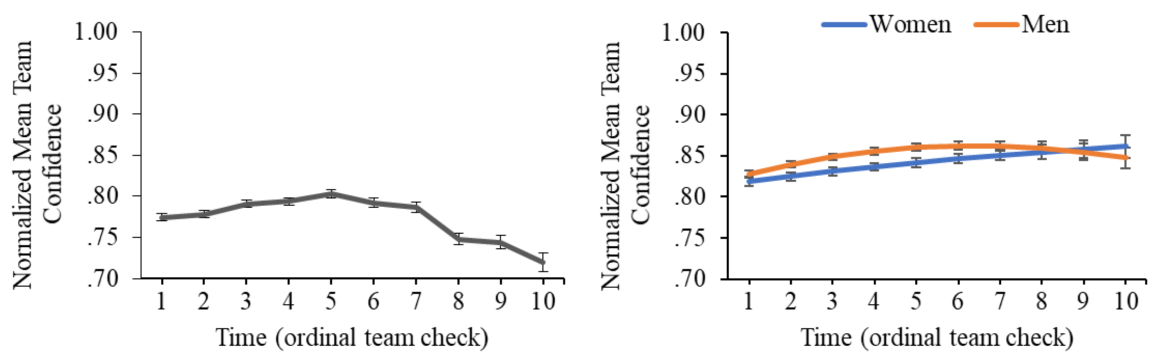

Figure 1. Normalized mean team confidence as a function of time (i.e., ordinal team check). The left panel shows the overall raw growth trajectory aggregating over gender, course, and Likert scale from an unstructured model (i.e., not accounting for clustering of responses in students within courses). The right panel shows the quadratic growth trajectories for women and men from a random effects model, with random intercepts and time slopes for both students and courses.

Analysis 2: The collective efficacy ratings are no different comparing midterm to end-of-term

Overall, collective efficacy ratings at mid-term ($M = 69.55$, $SE = .51$) were not significantly different from those at the end of the term ($M = 70.10$, $SE = .55$), $F(1, 613) = 1.35$, $p = .246$, $pseudo\text{-}R^2 = 1.71\%$, nor was there an interaction with gender, $F(1, 615) = 2.12$, $p = .146$, $pseudo\text{-}R^2 = 0.56\%$, or course, $F(2, 659) < 1$, $p = .502$, $pseudo\text{-}R^2 = 1.41\%$. The three-way interaction also was not significant, $F(2, 687) < 1$, $p = .839$, $pseudo\text{-}R^2 = 0.09\%$. No other effects were significant (*p*s > .15).

Analysis 3: Following the imposter syndrome lesson, team confidence ratings increased for men but not women

Overall, the team confidence ratings preceding the imposter syndrome lesson ($M = 7.41$, $SE = .127$) were not significantly different from the post-lesson ($M = 7.42$, $SE = .056$), $F(1, 904) < 1$, $p = .994$, $pseudo\text{-}R^2 = 0.08\%$. However, as shown in Figure 2, there was a significant interaction of pre-/post-lesson ratings and gender, $F(1, 950) = 5.45$, $p = .019$, $pseudo\text{-}R^2 = 4.95\%$, such that team confidence increased following the lesson for men ($b = .299$, $SE = .079$, $p < .001$) whereas team confidence was unaffected for women ($b = .001$, $SE = .100$, $p = .994$). This pattern was consistent across courses (engineering, business, and others), $F(2, 749) = 1.18$, $p = .307$, $pseudo\text{-}R^2 = 3.70\%$, and did not depend on the team confidence Likert scale having five versus nine points, $F(1, 980) = 1.16$, $p = .281$, $pseudo\text{-}R^2 = 0.13\%$. No other effects were significant (*p*s > .11).

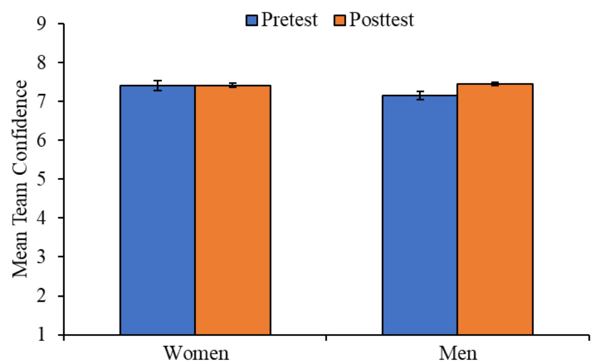

Figure 2. Mean team confidence ratings before (pretest) and after (posttest) the imposter syndrome lesson for women and men. Error bars represent +/- 1 standard error.

Discussion and implications

In these analyses, we find students reporting fairly stable team confidence and collective efficacy over the course of their semester projects, results that seem to hold both in first-year engineering design courses and other courses at our institution. While both men and women showed this general pattern, we find that women's ratings increase over the course whereas men's increase more quickly, then peak and fall back off. We also find that a tailored lesson on imposter syndrome is related to an increase in men's ratings of team confidence, but not women's. Though differences in the team confidence trajectory (analysis 1) and an interaction with our lesson intervention (analysis 3) are statistically significant, the effect sizes on these results are small; overall, both men and women generally report feeling fairly high team-referenced efficacy. As compared with prior work in STEM showing more stark self-efficacy differences between men and women, the team-based pedagogies in use across these Tandem courses may be an attenuating factor. Indeed, courses that typically employ high-stakes exams show disparate outcomes for female students in terms of both self-efficacy and grades [52], [53].

Especially with respect to the observed differences in trajectory of team confidence, these findings are not particularly surprising in the context of related work on gender and self-efficacy. Hirschfield and Chachra [54], for example, recently reported a study of first-year engineering design students very similar to the engineering population of interest here showing that men had higher confidence and self-efficacy scores in the beginning of the term but that women closed this gap by the end of the course. Marra and colleagues [55] reported a similar result for women in engineering programs. Although this study was conducted over the broader timescale of a full academic year, these researchers still found women increasing in terms of self-efficacy despite a significant decrease in feelings of inclusion. Overall, we have identified men and women patterning in fairly similar ways regarding confidence in team success.

As stated, the effect sizes for the gendered patterns we have identified are relatively small, nonetheless we highlight some potential pedagogical implications. Just as self-efficacy is malleable and subject to the effects of targeted interventions, researchers have shown that collective efficacy can be positively affected by short educational sessions on group work and cooperative learning [56], [57]. Importantly, these interventions were directed both at students and instructors. Our team has a similar goal to develop more and especially tailored Tandem content for instructors, partly in recognition that we must keep a critical eye on the extent to which we seek to support (change) students versus changing the learning environments to which they are exposed.

Developing Tandem content (such as a lesson) on the relationship between confidence and ability may be a useful pathway for mitigating the small gendered patterns we identified here. The Dunning-Kruger effect [58], for example, describes the phenomenon of being ignorant of one's own ignorance. With the goal of mitigating overconfidence (particularly for men and perhaps especially following the imposter syndrome lesson), a lesson on the Dunning-Kruger effect might help address gendered differences in team confidence and collective efficacy. Increasing students' awareness of these concepts may act as an important mediator of team success [59]. In future work we will also relate student ratings of team confidence and collective efficacy with external metrics of teamwork success.

Acknowledgements

This work would not have been possible without the students who have used Tandem and our faculty course partners. Robin R. Fowler, Laura K. Alford, and Stephanie A. Sheffield are the faculty leads responsible for the conceptual foundation and inspiration of Tandem, while design and development have been led by the Center for Academic Innovation, in particular the software development and behavioral science teams. Related exploratory data analyses were conducted by Albert M. Lee, a former CAI graduate student fellow and Erin Murray supported data acquisition. Funding was provided by the College of Engineering Educational Innovation Accelerator program and the National Science Foundation (DUE 2120252). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[2] C. M. Vogt, D. Hocevar, and L. S. Hagedorn, "A Social Cognitive Construct Validation: Determining Women's and Men's Success in Engineering Programs," *The Journal of Higher Education*, vol. 78, no. 3, pp. 337–364, May 2007, doi: 10.1353/jhe.2007.0019.

[3] A. Bandura, *Self-Efficacy: The Exercise of Control*. Macmillan, 1997.

[4] F. Pajares, “Self-Efficacy Beliefs in Academic Settings,” *Review of Educational Research*, vol. 66, no. 4, pp. 543–578, Dec. 1996, doi: 10.3102/00346543066004543.
[5] P. R. Pintrich and E. V. de Groot, “Motivational and self-regulated learning components of classroom academic performance,” *Journal of Educational Psychology*, vol. 82, no. 1, pp. 33–40, 1990, doi: 10.1037/0022-0663.82.1.33.
[6] G. V. Caprara, M. Vecchione, G. Alessandri, M. Gerbino, and C. Barbaranelli, “The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study,” *British Journal of Educational Psychology*, vol. 81, no. 1, pp. 78–96, 2011, doi: 10.1348/2044-8279.002004.
[7] K. Talsma, B. Schüz, R. Schwarzer, and K. Norris, “I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance,” *Learning and Individual Differences*, vol. 61, pp. 136–150, Jan. 2018, doi: 10.1016/j.lindif.2017.11.015.
[8] L. Stankov, J. Lee, W. Luo, and D. J. Hogan, “Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety?,” *Learning and Individual Differences*, vol. 22, no. 6, pp. 747–758, Dec. 2012, doi: 10.1016/j.lindif.2012.05.013.
[9] T. Nielsen, J. Dammeyer, M. L. Vang, and G. Makransky, “Gender Fairness in Self-Efficacy? A Rasch-Based Validity Study of the General Academic Self-Efficacy Scale (GASE),” *Scandinavian Journal of Educational Research*, vol. 62, no. 5, pp. 664–681, Sep. 2018, doi: 10.1080/00313831.2017.1306796.
[10] G. M. D’Lima, A. Winsler, and A. Kitsantas, “Ethnic and Gender Differences in First-Year College Students’ Goal Orientation, Self-Efficacy, and Extrinsic and Intrinsic Motivation,” *The Journal of Educational Research*, vol. 107, no. 5, pp. 341–356, Sep. 2014, doi: 10.1080/00220671.2013.823366.
[11] C. Huang, “Gender differences in academic self-efficacy: a meta-analysis,” *Eur J Psychol Educ*, vol. 28, no. 1, pp. 1–35, Mar. 2013, doi: 10.1007/s10212-011-0097-y.
[12] I. Johnson, “Female Faculty Role Models, Self-Efficacy and Student Achievement,” *College Student Journal*, vol. 51, no. 1, pp. 151–172, Mar. 2017.
[13] S. Pina-Neves, L. Faria, and H. Räty, “Students’ individual and collective efficacy: joining together two sets of beliefs for understanding academic achievement,” *Eur J Psychol Educ*, vol. 28, no. 2, pp. 453–474, Jun. 2013, doi: 10.1007/s10212-012-0123-8.
[14] R. L. Matz *et al.*, “Patterns of Gendered Performance Differences in Large Introductory Courses at Five Research Universities,” *AERA Open*, vol. 3, no. 4, p. 2332858417743754, Oct. 2017, doi: 10.1177/2332858417743754.
[15] E. Seymour and A.-B. Hunter, *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*. Springer International Publishing, 2019.
[16] D. L. Reinholz, R. L. Matz, R. Cole, and N. Apkarian, “STEM Is Not a Monolith: A Preliminary Analysis of Variations in STEM Disciplinary Cultures and Implications for Change,” *LSE*, vol. 18, no. 4, p. mr4, Dec. 2019, doi: 10.1187/cbe.19-02-0038.
[17] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, “Why are some STEM fields more gender balanced than others?,” *Psychological Bulletin*, vol. 143, no. 1, pp. 1–35, Jan. 2017, doi: http://dx.doi.org/10.1037/bul0000052.
[18] Michigan Engineering, “Facts & figures,” *Michigan Engineering*. https://www.engin.umich.edu/about/facts-figures/ (accessed Apr. 12, 2022).
[19] National Center for Education Statistics, “Digest of Education Statistics, 2019,” 2019. https://nces.ed.gov/programs/digest/d19/tables/dt19_322.50.asp (accessed Jan. 17, 2021).

[20] M. Keough, R. Fowler, and L. Hirshfield, “How Male Students Talk About the Female Student Experience on Teams,” presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Apr. 12, 2022. [Online]. Available: http://peer.asee.org/how-male-students-talk-about-the-female-student-experience-on-teams
[21] S. G. Brainard and L. Carlin, “A Six-Year Longitudinal Study of Undergraduate Women in Engineering and Science*,” *Journal of Engineering Education*, vol. 87, no. 4, pp. 369–375, 1998, doi: 10.1002/j.2168-9830.1998.tb00367.x.
[22] J. McKinney, M.-L. Chang, and D. Glassmeyer, “Why Females Choose STEM Majors: Understanding the Relationships Between Major, Personality, Interests, Self-Efficacy, and Anxiety,” *Journal for STEM Educ Res*, vol. 4, no. 3, pp. 278–300, Dec. 2021, doi: 10.1007/s41979-021-00050-6.
[23] U. Tellhed, M. Bäckström, and F. Björklund, “Will I Fit in and Do Well? The Importance of Social Belongingness and Self-Efficacy for Explaining Gender Differences in Interest in STEM and HEED Majors,” *Sex Roles*, vol. 77, no. 1, pp. 86–96, Jul. 2017, doi: 10.1007/s11199-016-0694-y.
[24] C. R. Fisher, C. D. Thompson, and R. H. Brookes, “Gender differences in the Australian undergraduate STEM student experience: a systematic review,” *Higher Education Research & Development*, vol. 39, no. 6, pp. 1155–1168, Sep. 2020, doi: 10.1080/07294360.2020.1721441.
[25] A. M. L. Cavallo, W. H. Potter, and M. Rozman, “Gender Differences in Learning Constructs, Shifts in Learning Constructs, and Their Relationship to Course Achievement in a Structured Inquiry, Yearlong College Physics Course for Life Science Majors,” *School Science and Mathematics*, vol. 104, no. 6, pp. 288–300, 2004, doi: 10.1111/j.1949-8594.2004.tb18000.x.
[26] C. J. Atman *et al.*, “Enabling Engineering Student Success: The Final Report for the Center for the Advancement of Engineering Education,” Morgan & Claypool Publishers, San Rafael, CA, 2010.
[27] E. Litzler, C. C. Samuelson, and J. A. Lorah, “Breaking it Down: Engineering Student STEM Confidence at the Intersection of Race/Ethnicity and Gender,” *Research in Higher Education*, vol. 55, no. 8, pp. 810–832, 2014.
[28] J. A. Ross, G. Scott, and C. D. Bruce, “The Gender Confidence Gap in Fractions Knowledge: Gender Differences in Student Belief–Achievement Relationships,” *School Science and Mathematics*, vol. 112, no. 5, pp. 278–288, 2012, doi: 10.1111/j.1949-8594.2012.00144.x.
[29] D. Chachra, A. Dillon, E. Spingola, and B. Saul, “Self-efficacy and task orientation in first-year engineering design courses,” in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, Oct. 2014, pp. 1–4. doi: 10.1109/FIE.2014.7044007.
[30] N. A. Mamaril, E. L. Usher, C. R. Li, D. R. Economy, and M. S. Kennedy, “Measuring Undergraduate Students’ Engineering Self-Efficacy: A Validation Study,” *Journal of Engineering Education*, vol. 105, no. 2, pp. 366–395, 2016, doi: 10.1002/jee.20121.
[31] O. Eris *et al.*, “A Preliminary Analysis Of Correlates Of Engineering Persistence: Results From A Longitudinal Study,” Jun. 2007, p. 12.94.1-12.94.12. Accessed: Feb. 18, 2022. [Online]. Available: http://peer.asee.org/a-preliminary-analysis-of-correlates-of-engineering-persistence-results-from-a-longitudinal-study
[32] D. Bressoud, V. Mesa, and C. Rasmussen, “Insights and recommendations from the MAA

national study of college calculus (MAA notes),” The Mathematical Association of America, Washington, DC, 2015.
[33] A. Bandura, “Exercise of Human Agency Through Collective Efficacy,” *Curr Dir Psychol Sci*, vol. 9, no. 3, pp. 75–78, Jun. 2000, doi: 10.1111/1467-8721.00064.
[34] K. Tasa, S. Taggar, and G. H. Seijts, “The development of collective efficacy in teams: A multilevel and longitudinal perspective,” *Journal of Applied Psychology*, vol. 92, no. 1, pp. 17–27, Jan. 2007, doi: 10.1037/0021-9010.92.1.17.
[35] A. D. Stajkovic, D. Lee, and A. J. Nyberg, “Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and test of a mediation model,” *Journal of Applied Psychology*, vol. 94, no. 3, pp. 814–828, 2009, doi: 10.1037/a0015659.
[36] A. Bandura, “Perceived Self-Efficacy in Cognitive Development and Functioning,” *Educational Psychologist*, vol. 28, no. 2, pp. 117–148, Mar. 1993, doi: 10.1207/s15326985ep2802_3.
[37] J. Z. N. Khong, G. A. D. Liem, and R. M. Klassen, “Task performance in small group settings: the role of group members’ self-efficacy and collective efficacy and group’s characteristics,” *Educational Psychology*, vol. 37, no. 9, pp. 1082–1105, Oct. 2017, doi: 10.1080/01443410.2017.1342767.
[38] D. A. Shearer, P. Holmes, and S. D. Mellalieu, “Collective efficacy in sport: the future from a social neuroscience perspective,” *International Review of Sport and Exercise Psychology*, vol. 2, no. 1, pp. 38–53, Mar. 2009, doi: 10.1080/17509840802695816.
[39] E. Yaakobi and J. Weisberg, “Organizational Citizenship Behavior Predicts Quality, Creativity, and Efficiency Performance: The Roles of Occupational and Collective Efficacies,” *Frontiers in Psychology*, vol. 11, 2020, Accessed: Feb. 18, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2020.00758
[40] R. J. Sampson, S. W. Raudenbush, and F. Earls, “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy,” *Science*, vol. 277, no. 5328, pp. 918–924, Aug. 1997, doi: 10.1126/science.277.5328.918.
[41] D. Maimon and C. R. Browning, “Unstructured Socializing, Collective Efficacy, and Violent Behavior Among Urban Youth*,” *Criminology*, vol. 48, no. 2, pp. 443–474, 2010, doi: 10.1111/j.1745-9125.2010.00192.x.
[42] A. A. Niler, R. Asencio, and L. A. DeChurch, “Solidarity in STEM: How Gender Composition Affects Women’s Experience in Work Teams,” *Sex Roles*, vol. 82, no. 3, pp. 142–154, Feb. 2020, doi: 10.1007/s11199-019-01046-8.
[43] K. Forslund Frykedal, E. H. Chiriac, and M. Rosander, “Efficacy beliefs and interdependence when being assessed working in a group,” *Educational Studies*, vol. 47, no. 5, pp. 509–520, Sep. 2021, doi: 10.1080/03055698.2019.1706039.
[44] C.-C. (Timothy) Lin and T.-K. (T K.) Peng, “From Organizational Citizenship Behaviour to Team Performance: The Mediation of Group Cohesion and Collective Efficacy,” *Management and Organization Review*, vol. 6, no. 1, pp. 55–75, Mar. 2010, doi: 10.1111/j.1740-8784.2009.00172.x.
[45] R. Fowler, L. K. Alford, S. Sheffield, C. Hayward, T. S. Henderson, and R. L. Matz, “Supporting Equitable Team Experiences Using Tandem, an Online Assessment and Learning Tool,” presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Apr. 12, 2022. [Online]. Available: http://peer.asee.org/supporting-equitable-team-experiences-using-tandem-an-online-assessment-and-learning-tool

[46] W. R. Miller and S. Rollnick, *Motivational Interviewing: Helping People Change*. New York: Guilford Press, 2012.
[47] President's Council of Advisors on Science and Technology, "Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics," Washington, DC, 2012. Accessed: Aug. 14, 2013. [Online]. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf
[48] R. L. Matz, A. M. Lee, R. R. Fowler, and C. Hayward, "Teammates Stabilize over Time in How They Evaluate Their Team Experiences," in *LAK22: 12th International Learning Analytics and Knowledge Conference*, New York, NY, USA, Mar. 2022, pp. 485–491. doi: 10.1145/3506860.3506891.
[49] American Psychological Association, "Gender," *https://apastyle.apa.org*, Sep. 2019. https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/gender (accessed Feb. 18, 2022).
[50] R. W. Lent, J. Schmidt, and L. Schmidt, "Collective efficacy beliefs in student work teams: Relation to self-efficacy, cohesion, and performance," *Journal of Vocational Behavior*, vol. 68, no. 1, pp. 73–84, Feb. 2006, doi: 10.1016/j.jvb.2005.04.001.
[51] P. R. Clance and S. A. Imes, "The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention," *Psychotherapy: Theory, Research & Practice*, vol. 15, no. 3, pp. 241–247, 1978, doi: 10.1037/h0086006.
[52] S. Salehi *et al.*, "Gender Performance Gaps Across Different Assessment Methods and the Underlying Mechanisms: The Case of Incoming Preparation and Test Anxiety," *Frontiers in Education*, vol. 4, 2019, Accessed: Feb. 18, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/feduc.2019.00107
[53] E. M. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, "Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?," *Phys. Rev. Phys. Educ. Res.*, vol. 14, no. 2, p. 020123, Dec. 2018, doi: 10.1103/PhysRevPhysEducRes.14.020123.
[54] L. J. Hirshfield and D. Chachra, "Complex Systems: Interactions Between Gender and Project Context on Confidence of First-Year Engineering Students," *International Journal of Engineering Education*, vol. 37, no. 4, pp. 1090–1107, 2021.
[55] R. M. Marra, K. A. Rodgers, D. Shen, and B. Bogue, "Women Engineering Students and Self-Efficacy: A Multi-Year, Multi-Institution Study of Women Engineering Student Self-Efficacy," *Journal of Engineering Education*, vol. 98, no. 1, pp. 27–38, 2009, doi: 10.1002/j.2168-9830.2009.tb01003.x.
[56] E. H. Chiriac, M. Rosander, and K. F. Frykedal, "An Educational Intervention to Increase Efficacy and Interdependence in Group Work," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3415260, Jun. 2019. Accessed: Feb. 18, 2022. [Online]. Available: http://papers.ssrn.com/abstract=3415260
[57] M. Rosander, K. Forslund Frykedal, and E. Hammar Chiriac, "Attitudes towards being assessed in group work: The effects of self-efficacy and collective efficacy moderated by a short educational intervention," *Psychology in the Schools*, vol. 57, no. 9, pp. 1404–1416, 2020, doi: 10.1002/pits.22423.
[58] D. Dunning, "Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance," in *Advances in Experimental Social Psychology*, vol. 44, J. M. Olson and M. P. Zanna, Eds. Academic Press, 2011, pp. 247–296. doi:

10.1016/B978-0-12-385522-0.00005-6.
[59] C. M. Kiesel, “Perceptions of Collective Efficacy as a Mediator: An Examination of the Perceptions of Group Cohesion, Social Loafing, and Collective Efficacy,” Xavier University, 2018. Accessed: Feb. 18, 2022. [Online]. Available: http://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=xavier1534937246305126